JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc

Article

Specifications-Based Grading Facilitates Student−Instructor Interactions in a Flipped-Format General Chemistry II Course

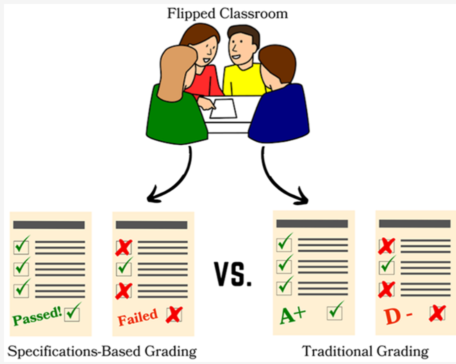
Broden Bunnell, Lauren LeBourgeois, James Doble, Brian Gute, and Jacob W. Wainman*

Cite This: *J. Chem. Educ.* 2023, 100, 4318−4326

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: General chemistry II is often a challenging course for first-year STEM students, and many students rely on partial credit to achieve success in the course. However, assigning liberal partial credit to students' work can also disincentivize developing proficiency with the material. Here, we present an alternative specifications-based grading scheme for a general chemistry II lecture delivered as a flipped class. Students evaluated with a specifications-based grading scheme had higher average end-of-semester grades, with a greater proportion of students receiving A's and A−'s than the students in the traditional grading scheme. Performance on standardized ACS final exams decreased slightly in courses using specifications-based grading. Additionally, we present the results of end-of-semester surveys, which revealed that students under the specifications-based grading system agreed more with positive statements related to student-instructor interactions, including getting feedback and asking for help. These results suggest that specifications-based grading may lead to more positive outcomes in the course and may encourage students to interact with their instructor more often.



KEYWORDS: *First-Year Undergraduate/General, Curriculum, Collaborative/Cooperative Learning, Student-Centered Learning*

■ INTRODUCTION

General chemistry is a common series of courses for college science, technology, engineering, and mathematics (STEM) students, as robustly understanding the properties of matter is paramount to success in STEM professions. The second semester of general chemistry (referred to hereafter as general chemistry II) includes concepts such as thermodynamics, kinetics, and equilibrium and involves mathematical problem-solving skills that are challenging for many first-year college students. Thus, general chemistry II represents a significant barrier to persistence in STEM, particularly for students from historically underserved populations.[1,2] Despite this challenge, general chemistry II topics are broadly applicable to both real STEM problems and in subsequent STEM courses within and outside of chemistry.[3] It is vital that students are proficient with the course content to be successful in STEM coursework and careers.

Traditionally, student work is graded using point-based grading systems where partial credit is awarded. Specifications-based grading, in contrast, does not use points or partial credit; instead, students earn credit by demonstrating their proficiency with the course material on assessments (see Table 1).[4] Where students in traditional grading schemes earn points, specifications-based grading opts for a "pass/fail" system that requires students to meet a minimum standard of proficiency. In traditional grading, students often do not revisit their work to master concepts that received partial credit, whereas specifications-based grading offers students second chances to demonstrate their proficiency with the material.[5]

Specifications-based grading has been implemented in several chemistry courses[1,5−17] and other STEM fields.[18−24] Noell et al. created a hybrid specifications-based grading system for small class sections of general chemistry II and found more final grades of A's compared to traditional grading.[6] Martin[5] and Toledo and Dubas[8] introduced variations of specifications-based grading into general chemistry I, showing notable positive student outcomes including higher final exam scores (Martin) and higher levels of understanding of the final exam compared to unit exams (Toledo and Dubas). Hollinsed used specifications-based grading in general chemistry and found an increase in the amount of A's and no effect on the DFW rates (i.e., the proportion of students earning a D or F or withdrawing from the course).[17] Motivated by these data, we implemented specifications-based grading in a flipped-format general chemistry II course. Previous studies have shown that "flipped" courses can reduce DFW rates and improve student perform-

Received: May 20, 2023
Revised: October 18, 2023
Accepted: October 18, 2023
Published: October 31, 2023

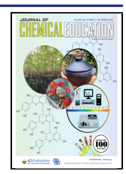


Table 1. Traditional vs. Specifications-Based Grading

	traditional grading	specifications-based grading
in-semester grading	points-based grading system	proficiency "pass/fail" grading system
	students earn partial credit determined by the instructor	no partial credit is awarded, and students earn a passing grade when they achieve high scores (e.g., 80% correct)
	only single attempts are allowed on the assessments and exams	students have second chances to earn passing grades on the assessments and exams
final grades	determined based on the percent of the total points earned during the semester	determined based on the amount of assignments and exams where a student successfully demonstrated proficiency

Table 2. Assessment Strategy

assessment	formative or summative?	format	specifications for demonstrating proficiency	rationale for specifications
pre-class work	formative	individual, online quiz	no more than two questions wrong on the quiz, completed on time	proficiency with the material was not expected before class.
in-class work	formative	small group worksheet	all work shown, completed and submitted on time	worksheets were designed to develop proficiency, so grading on completion was sufficient
homework	formative/ summative	individual, online	score of 85% or greater, submitted on time	higher scores were expected since these were completed after in-class practice
exams	summative	individual, in-person, two attempts	satisfactory performance (i.e., correct answers) on 80% or more of the exam questions	students were expected to reach at least B− level work after two attempts on the exams, and feedback was given after the first attempt
final exam	summative	ACS general chemistry II standardized exam	earning a score of 60% or greater	this score was considered acceptable proficiency due to the ACS exam's rigor and cumulative breadth and because the multiple-choice format was unlike other exams in the course

ance outcomes.[25,26] Our work fills a gap in the literature by applying specifications-based grading to a "flipped" general chemistry II course.

Here, we present our specifications-based grading scheme, which requires students to demonstrate proficiency on pre-class work, in-class work, homework, and exams. We describe our grading model, assessments, implementation tips, and lessons learned. We also analyzed student outcomes and survey data that aimed to address the following research questions: (1) how did the use of specifications-based grading impact students' performance in the course and on final exams, (2) did this impact vary with gender, and (3) how did the grading scheme influence students' perceptions of the "flipped" course format?

■ COURSE FORMAT

This general chemistry II course was the second course in a two-semester college-level chemistry sequence required for STEM majors and students pursuing professional programs (e.g., medical school). At our institution, the laboratory component is an independent course falling under a separate course number, often with a different instructor (see Supporting Information 1 for institutional details). All of the courses included in this study were delivered in a "flipped" format, which is described elsewhere.[27] This format asked students to complete pre-class videos and quizzes, in-class group worksheets, online homework, and exams. In the traditionally graded course, these assignments were all assigned points, and the problem-based questions were graded using partial credit. See Supporting Information 1 for course format details.

■ SPECIFICATIONS-BASED ASSESSMENTS

Student learning outcomes were assessed following a specifications-based grading scheme. In keeping with specifications-based grading, the minimum criteria for "demonstrating proficiency" on each of the assessment types (i.e., pre-class work, in-class work, homework, and exams) were clearly defined. See Table 2 below for a summary of the assessment types and specifications for demonstrating proficiency. See Supporting Information 1 for assessment details.

Pre-Class Work: Individual Online Quizzes

Due to the "flipped" format of the course, students watched pre-class videos and/or did pre-class reading from the textbook. Then, the students completed short, multiple-choice quizzes (typically six questions each) on the material covered in these videos or readings. Students were allowed only one attempt but were able to use any resources while completing these quizzes, including their notes, the textbook, and online resources. The criteria for "demonstrating proficiency" for each quiz was (1) completing the quiz before the deadline (30 min before class) and (2) no more than two incorrect answers on the quiz. This low score was deemed sufficient proficiency because these quizzes were formative assessments and because students only had one attempt; the subsequent class period was used to develop deeper understanding.

In-Class Work: Group Worksheets

During each 75 min class period, students worked in small groups completing worksheets provided by the instructor. These worksheets emphasized quantitative problem solving in line with the learning objectives of general chemistry II (see Supporting Information 3 for an example worksheet). To grade the in-class worksheets, criteria for "demonstrating proficiency" were set as (1) completing and submitting the worksheet on time (i.e., 2−3 days after the worksheet was assigned) and (2) showing all work used to solve for the correct answers. Using completion as a criterion for demonstrating proficiency solved two problems. First, worksheet completion was incentivized, which encouraged student groups to solve all the problems. The resulting document could then be used by group members during exam review. Second, grading student work by completion lessened the grading load of daily worksheets.

Homework: Individual Online Assignments

Following each class period, students were assigned 5−7 homework problems administered through an online home-

https://doi.org/10.1021/acs.jchemed.3c00473
J. Chem. Educ. 2023, 100, 4318−4326

work system. The criteria for "demonstrating proficiency" for each assignment was (1) completing the assignment on time (i.e., 1 week after the class period covering the corresponding content) and (2) earning at least 85% of the points on the assignment while having unlimited attempts on each problem. This relatively high score (85%) was chosen to represent a sufficient demonstration of proficiency because these assignments were meant to encourage students to practice and review problems aligned with the learning objectives from the pre-class and in-class work.

Exams: Individual Tests with Multiple Attempts

Students completed four, 1 h regular exams throughout the semester and one, 2 h cumulative final exam. Each regular exam consisted of ten problems (sometimes organized as five two-part problems). Most of these problems had quantitative solutions, in line with the learning objectives of general chemistry II, but some were more qualitative or conceptual in nature. Students' answers to each of these ten problems were evaluated on a trinary scale: S, S−, or N. Answers earning a "satisfactory" rating (S) were fully correct, having the correct units and the correct number of significant digits. Incorrect student answers were marked as "not satisfactory" (N). The "somewhat satisfactory" rating (S−) was assigned to answers that contained a small and easy-to-identify error (e.g., a decimal transposition or an incorrect number of significant digits). Simple errors that reflected an incorrect understanding of the underlying chemical concepts (e.g., reversing the sign of a thermodynamic quantity like entropy) or that resulted from common algebraic errors earned an "N".

The criteria for "demonstrating proficiency" on the entire exam was earning S ratings on eight out of the ten exam questions (i.e., B− work). S− ratings on two exam questions were considered equivalent to one S rating. For example, a student earning seven S ratings, two S− ratings, and one N rating was still considered as having demonstrated sufficient proficiency on the exam. In addition, we considered exams with S ratings on only seven out of ten problems as a "somewhat satisfactory" demonstration of proficiency, akin to the S− ratings on the individual exam problems. Those unfamiliar with specifications-based grading may find that earning a B− on exams is a high standard that, under a traditional grading scheme, is difficult to achieve on high-stakes assessments with no partial credit. In fact, the percentage of students demonstrating proficiency on the exam was low, varying from 10−30% depending on the exam material. This result was addressed following another tenet of specifications-based grading; students had a second attempt on each exam to demonstrate their proficiency.

Prior to the second attempt on each exam, the students' first attempt exams were graded, and students were provided with feedback on which answers had earned S, S−, or N ratings. In practice, this notification was achieved using cover sheets on the first attempt exam (see Supporting Information 3 for an example). These cover sheets contained ten blanks where students wrote their final answers to the ten exam questions. The blanks were labeled with the concepts tested by each question and the quantity being calculated in each problem. Each problem was scored as earning S, S−, or N on these cover sheets before being handed back. Because the cover sheets did not contain the exam questions or the students' work from the first exam attempt, students were encouraged to use the labels on the cover sheet to determine what problems to review prior to the second attempt. Also, we documented common errors leading to S− or N scores while grading the exams, including an explanation of the error or misconception leading to the common incorrect answers. We disseminated a document containing these errors and descriptions for students to review. Additionally, students were allowed to review their worked out exams during faculty office hours upon request.

Approximately 1 week after handing back the cover sheets, students could then retry any exam problems for which they did not earn an S rating. Thus, students had ample opportunity in the week between receiving their cover sheets and taking their second attempt exam to review their mistakes, visit faculty and TA office hours, rewatch pre-class videos, and practice similar problems from the in-class worksheets and textbook.

The second attempt exams were scheduled outside of class on dates provided in the course syllabus (see Supporting Information 1 for second attempt exam administration details). The second attempt exams consisted of isomorphic problems (problems that were similar but not identical to the problems on the first attempt exam) to prevent students from simply memorizing the correct answers from the first attempt. On the second attempt exams, students completed only the problems from the first attempt that had received an N or S− rating. Upon completing the second exam attempts, the percent of students earning an S on the exams increased to 46−64%, again varying based on the difficulty of the exam material. These increased exam performances reveal that student outcomes were improved.

Final Exam: ACS Standardized General Chemistry II Exam

For the purposes of semester-by-semester comparison, the American Chemical Society's general chemistry II exam was used as the final summative assessment for the course. To fit the scores from this standardized multiple-choice exam into our specifications-based grading scheme, the criteria for "demonstrating proficiency" on the final exam was set at achieving a score of 42 out of 70, equivalent to the 70th percentile. The criteria for "somewhat satisfactory" proficiency on the final exam was set at a score of 35 out of 70, equivalent to the 50th percentile. These relatively low percentile scores were chosen because the ACS exam's format (multiple choice) was unfamiliar to students. Moreover, there was concern that students who had grown accustomed to the low-stakes nature of exams could be disadvantaged by this end-of-semester change in assessment strategy.

Assigning Final Course Grades

Students' overall grades were assigned based on the number of assessments from each category (pre-class work, in-class work, homework, and exams) where the student sufficiently demonstrated proficiency. This information was presented to students in the course syllabus, as shown in Table 3.

To determine a student's final letter grade, we first counted the number of exams where each student successfully achieved the criteria for demonstrating proficiency (including the final exam). We also considered the combination of two "somewhat satisfactory" exams to be equivalent to one "satisfactory" exam.

For pre-class work, in-class work, and homework, there were 25 assignments per category throughout the semester. So, the number of satisfactory assignments from each of these non-exam assessment categories was used to select which column of the table contained their final grade. The numbers labeling the columns of Table 3 denote the minimum numbers of

Table 3. Determining Students' Final Letter Grades

number of satisfactory pre-class works, in-class works, or homeworks		22, 22, 22	22, 22, 18	22, 22, 14	22, 18, 18	22, 22, 10	22, 18, 14	18, 18, 18	22, 18, 10	22, 14, 14	18, 18, 14	22, 14, 10	18, 18, 10	18, 14, 14	22, 10, 10	18, 14, 10	14, 14, 14	18, 10, 10	14, 14, 10	14, 10, 10	10, 10, 10
number of satisfactory exams[a]	5	A	A–	B+	B+	B+	B+	B+	B+	B+	B+	B	B	B	B	B	B	B–	B–	C+	C+
	4.5	A–	A–	B+	B+	B+	B+	B+	B	B	B	B–	B–	B–	B–	B–	B–	B–	B–	C+	C+
	4	B+	B+	B	B	B	B	B	B–	B–	B–	C+	C+	C+	C+	C+	C+	C+	C+	C	C
	3.5	B+	B	B–	B–	B–	B–	B–	B–	B–	B–	C+	C+	C+	C+	C+	C	C	C–	C–	C–
	3	B	B–	C+	C+	C+	C+	C+	C+	C+	C+	C	C	C	C	C	C	C–	C–	D+	D+
	2.5	B–	B–	C+	C+	C+	C+	C+	C	C	C	C–	C–	C–	C–	C–	C–	C–	C–	D+	D+
	2	C+	C+	C	C	C	C	C	C–	C–	C–	D+	D+	D+	D+	D+	D+	D+	D+	D	D

[a]Note: The non-integer rows of the table are the result of "somewhat satisfactory" exams, which count as 0.5 satisfactory exams.

satisfactory assignments from each of these categories, assigned in any order.

To clearly illustrate how this table was used to assign final letter grades, consider a hypothetical example student, Jane. At the end of the semester, Jane had two satisfactory exams (8/10) and three somewhat satisfactory exams (7/10). Thus, we find her final grade in the fourth row of the table (2 + 3(0.5) = 3.5 satisfactory exams). Jane also successfully demonstrated proficiency on 23 pre-class works, 21 in-class works, and 24 homework assignments. Since the three numbers labeling each column can be applied to the non-exam assessments in any order, Jane has met the minimum specifications for the second column in the table labeled as 22, 22, 18 (24 homeworks > 22, 23 pre-class works > 22, 21 in-class works > 18). Therefore, Jane would earn a B. If Jane had successfully demonstrated proficiency on one more in-class work, she would have met the specifications described by the first column (22, 22, 22), which would have earned her a B+ instead. If we had not accepted somewhat satisfactory exams, then Jane's grade would have come from the last row of the table (2 satisfactory exams), meaning she would have earned a C+.

Failure to meet the minimum specifications in any of the four assessment categories resulted in failing the class. Therefore, anyone unable to demonstrate proficiency on at least two exams or 10 of each of the non-exam assessments earned a final grade of an F. This F grade was assigned even if performance in other categories was outstanding. For example, a student who demonstrated proficiency on all five of the exams and all of the pre-class and in-class work but then failed to complete at least ten of the online homework assignments to the minimum specifications would fail the course.

Nuts and Bolts: Implementation of Specifications-Based Grading

The reality of using this grading strategy to assess student performance posed new and unique logistical challenges. Similarly, students required support to ensure they were successfully demonstrating proficiency according to the terms of this scheme. We have included key implementation tips in Supporting Information 1. These tips include (1) the history of how Table 3 was developed, (2) a description of how we made "on the fly" adjustments to the specifications during our first implementation, (3) the role of aligning learning objectives to the assessments, and (4) a description how to administer the "exam cycle", which consists of the first attempt exam, grading, providing feedback, and the second attempt exam.

■ DATA COLLECTION AND ANALYSIS

Student Performance Outcomes

To assess our implementation of specifications-based grading, we conducted a retrospective study of outcomes and survey results. We analyzed the final letter grades and final exam scores for general chemistry II students from courses using the traditional and specifications-based grading schemes. We analyzed data from seven semesters of general chemistry II, four with traditional grading (fall 2015, spring 2016, fall 2016, and spring 2017) and three with specifications-based grading (fall 2017, spring 2018, and fall 2018). Results did not significantly vary semester-by-semester within a given grading scheme, so the data from each grading scheme were pooled (N = 820 and 511 for traditional and specifications-based grading, respectively). To compare final grades, we converted students' letter grades to a grade point using a 0.0−4.0 grade point scale.

https://doi.org/10.1021/acs.jchemed.3c00473
J. Chem. Educ. 2023, 100, 4318−4326

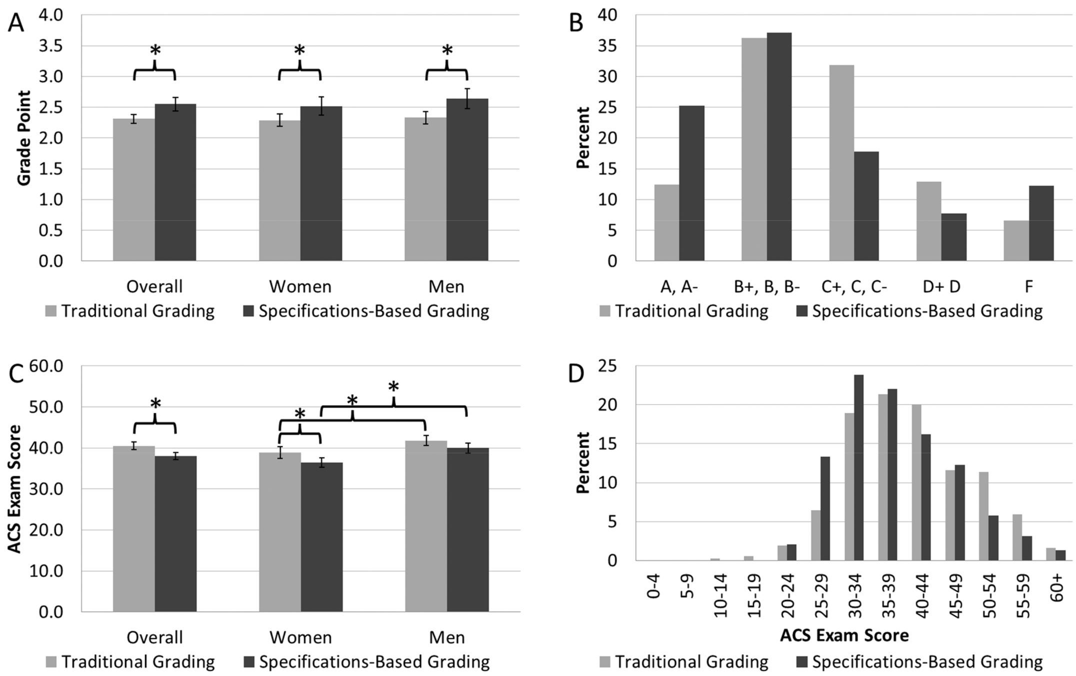


Figure 1. Student outcomes from implementing traditional and specifications-based grading in general chemistry II. Panel A shows increases in the average grade point earned by students graded by either a traditional point-based grading scheme (light gray bars) or a specifications-based grading scheme (dark gray bars). This increase was present overall or when disaggregated by gender. Error bars represent 95% confidence intervals. Stars represent significant differences between the data, indicated by braces by a Mann−Whitney U test. Panel B shows the distribution of letter grades for either grading scheme. Panel C shows decreases in ACS final exam scores in the course graded using a traditional point-based grading scheme vs a specifications-based grading scheme. The decrease was present overall or when disaggregated by gender. There are also differences in ACS exam results between men and women in both the traditionally graded course and the specifications-based graded course. Panel D shows the distribution of ACS exam scores for either grading scheme.

Final exam scores are the number of correct answers (out of 70). Statistical analyses were conducted in the Python package "scipy.stats". According to a Shapiro test, the data were not normally distributed. Thus, we used the Mann−Whitney U test to determine if differences between means were statistically significant. For family-wise comparisons in the disaggregated data, we used the Bonferroni Correction. Cohen's *d* was calculated to determine effect sizes. See Supporting Information 2 for the data analysis and statistical methods details.

Figure 1A shows a comparison of grade point averages (GPAs). The overall GPA from the pooled traditionally graded courses was 2.31 ± 0.07 (C+), while the overall GPA from the courses using specifications-based grading was 2.55 ± 0.11 (between C+ and B−). This increase was significant but small (effect size = 0.20). This result can be further analyzed by examining changes in the distributions of final grades (Figure 1B). The two distributions of grades were statistically significantly different. Using specifications-based grading, the proportion of students earning A's or A−'s increased, while the proportion of students earning C+'s, C's, C−'s, D+'s, and D's decreased, explaining the increase in GPA. Interestingly, the proportion of A's to A−'s shifted only slightly between the two courses; A's made up 53% and 65% of the total A's and A−'s for the traditionally graded and specifications-based graded courses, respectively. B+'s, B's, and B−'s remained relatively constant. The proportion of students earning F's did increase under our grading scheme, perhaps due to there being more pathways to earn an F (e.g., not achieving the minimum specifications on pre-class work, in-class work, homework, or exams shown in Table 3). Taken together, this resulted in a minor increase in the percent of students earning D's and F's from 19.5% in the traditionally graded course to 19.9% when using specifications-based grading. The proportion of students who withdrew from the course were similar at 6.7% and 7.7% for the traditional and specifications-based grading schemes, respectively.

Figure 1C shows a comparison of the final exam performance from both grading schemes. The average final exam score from the traditionally graded courses was 40.5 ± 0.9 (57.9%), while the average final exam score from the courses using specifications-based grading was 38.1 ± 0.9 (54.4%). This decrease was significant but small (effect size = 0.28). Both results remain higher than the national mean for this exam (36.2, 51.7%). In addition, we examined the distribution of final exam scores for both cohorts (Figure 1D). The two distributions of grades were not statistically significantly different.

To assess the impact of the grading schemes on students of various identities, we disaggregated the grade point averages and final exam scores by gender (Figures 1A and 1C, respectively). In the traditionally graded course, the GPA for women and men was 2.29 ± 0.10 (C+) and 2.33 ± 0.10 (C+), respectively, which are not statistically significantly different. Using specifications-based grading, the GPAs for both women and men increased to 2.52 ± 0.15 (between C+ and B−, effect size = 0.20) and 2.63 ± 0.16 (between C+ and B−, effect size = 0.26), respectively. The difference between women and men using specifications-based grading was not significant; however, the increase in GPAs observed for both genders between the two grading schemes was significant. The average final exam scores for women and men in the traditionally graded course

https://doi.org/10.1021/acs.jchemed.3c00473
J. Chem. Educ. 2023, 100, 4318−4326

were statistically significantly different at 38.8 ± 1.4 (59.7%) and 41.8 ± 1.2 (64.3%), respectively. In the specifications-based grading course, women had statistically significantly different average final exam scores of 36.4 ± 1.1 (56.0%, effect size = 0.27). Men's average final exam scores decreased to 40.0 ± 1.3 (61.5%), but this change is not significant.

Student Perception Results

To better understand how students' perceptions of the course format changed as a result of implementing this grading scheme, we conducted an end-of-semester survey. The survey contained several items related to students' attitudes toward chemistry as well as features of the course and the course instructor (Table 4). Students were asked to rank their level of agreement with each item on the survey on a four-point Likert scale (strongly agree, agree, disagree, strongly disagree). These rankings were converted to numerical values (4 = strongly agree to 1 = strongly disagree) for further analysis. The average rating for each survey item from students in either grading scheme is shown in Figure 2.

Table 4. End-of-Semester Survey Items

item number	item
1	I have a stronger interest in chemistry as a result of this course.
2	I have a better appreciation for the importance of chemistry as a result of this course.
3	Attending class enhances my learning in this course.
4	I get a lot of feedback from the instructor and/or from other students on how well I'm doing in this class.
5	It feels safe to volunteer an answer in this class, even if it is wrong.
6	I come to class early to talk with my classmates about assignments, homework, etc.
7	The instructor respects my opinion.
8	It is pretty obvious when someone has come to this class unprepared.
9	The instructor cares about whether I learn the course material.
10	I have studied for this course outside of the classroom with one or more of my classmates.
11	I feel comfortable asking for help from my classmates.
12	I can almost always get my questions answered by the instructor.
13	I feel comfortable asking for help from the instructor.
14	I've spoken informally with the instructor before, during, or after class.
15	In this class, I feel comfortable asking questions when I'm confused about something.
16	I am acquainted with the students sitting near me in class.
17	Before this semester, I was acquainted with the students sitting near me in class.
18	During class, I often have a chance to discuss material with some of my classmates.

Specifications-based grading led to students having a slightly higher interest in chemistry (item 1) and appreciation for the importance of chemistry (item 2). Students were more likely to agree that attendance in class was important (item 3). Also, students in courses using specifications-based grading were more likely to agree that they had positive interactions with the instructor. These interactions included getting feedback (item 4), being comfortable volunteering possible answers in class (item 5), and being comfortable asking questions (items 12 and 15) or asking for help (item 13). Students more strongly agreed that the instructor respected them (item 7) and cared about them learning the material (item 9) when using specifications-based grading. Students from the course using specifications-based grading were more likely to agree that they had met informally with the instructor (item 14). Interestingly, there was a significant decrease in students' ability to tell if a classmate was unprepared (item 8) when using the specifications-based grading scheme. These results suggest that a grading scheme based on demonstrating proficiency results in more positive attitudes toward student−instructor interactions. These results were independent of course instructor or course format (all courses in this study were "flipped" courses using identical course materials).

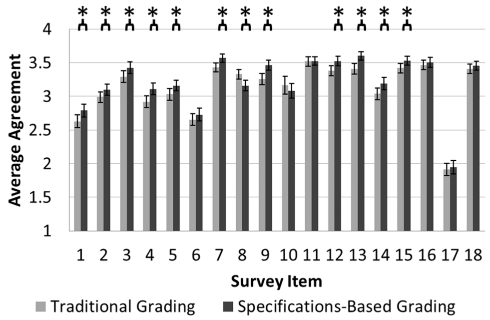


Figure 2. Student responses to an end-of-semester survey. Average survey responses are shown from courses using a traditional point-based grading scheme (light gray bars) or a specifications-based grading scheme (dark gray bars). Error bars represent 95% confidence intervals. Stars represent significant differences between the data indicated by braces by a Mann−Whitney U test.

These survey results were disaggregated by binary gender and compared between the two grading schemes (Figure 3). As seen in Figure 3A, there was increased agreement among women for many of the survey items, including many of the items that showed increases in the overall results described above (items 4, 5, 7, 9, and 12−15). In addition, women were more likely to agree that they came to class early (item 6). There was no longer a significant difference in women's interest (item 1) and appreciation (item 2) for chemistry. There was also no significant difference between women's perception of the importance of attending class (item 3) and their ability to tell if someone had come to class unprepared (item 8). In contrast, no significant increases in agreement with the survey items were observed among men in the courses using specifications-based grading (Figure 3B). There are qualitatively similar but not statistically significant increases in agreement for many but not all survey items among men.

Figures 3C and 3D rearrange the same data shown in Figures 3A and 3B for ease of comparison between men and women under the same grading scheme. In the traditionally graded course, we found that men were more likely than women to agree that the course gave them a stronger interest (item 1) and better appreciation (item 2) for chemistry. These gendered differences in items 1 and 2 were not present in the course using specifications-based grading. Using specifications-based grading, there were two significant differences in the survey responses between men and women. Women in these courses were more likely to agree that they came to class early (item 6) and that they had spoken informally with the instructor (item 14). Taken together, the survey results disaggregated by gender suggest that the improved instructor−student relationships under the new grading scheme revealed in the overall analysis were largely due to interactions between the instructors and women students. This result is important because faculty−student relationships are important

https://doi.org/10.1021/acs.jchemed.3c00473
J. Chem. Educ. 2023, 100, 4318−4326

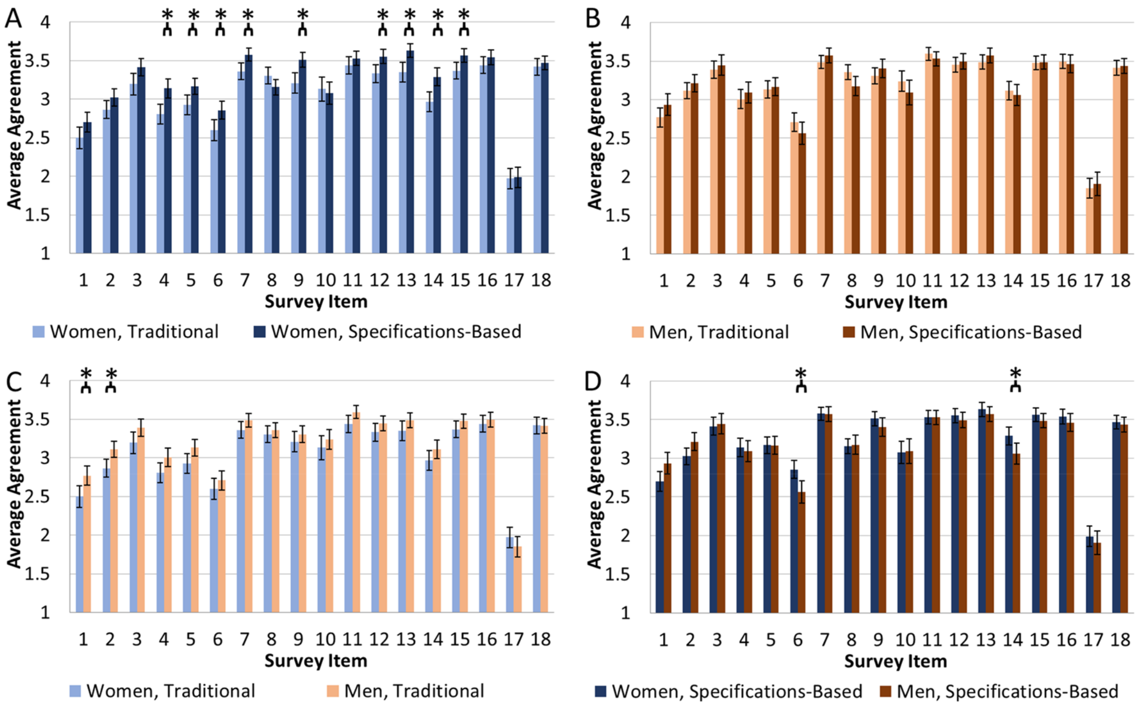


Figure 3. Student responses to an end-of-semester survey disaggregated by grading system and course type. Average survey responses are shown for women in the traditionally graded course (light blue bars), women under the specifications-based grading scheme (dark blue bars), men in the traditionally graded course (light orange bars), and men under the specifications-based grading scheme (dark orange bars). Error bars represent 95% confidence intervals. Stars represent significant differences between the data indicated by braces by a Mann−Whitney U test. Panels A−D show the same data organized for ease of pairwise comparisons.

for developing a sense of belonging in a classroom environment, and belonging is linked to persistence in STEM.[1,28−31]

DISCUSSION AND CONCLUSION

Overall, our results show that the implementation of specifications-based grading did improve students' final grades in general chemistry II, but it may not have improved learning according to the final exam results. The changes to the final grade distribution we observed mirror that of Noell et al.[6] and Hollinsed,[17] where the number of A's increased using specifications-based grading; however, we did not observe a similar increased withdrawal rate.[6] Toledo and Dubas,[8] in contrast, saw an increase in the number of B's, which remained constant in this study. Unlike our results, Martin saw increased performance on final exams.[5]

Using this grading scheme has similar impacts on the final grades of both women and men, leading to increases in their final letter grade of 0.23 and 0.30 grade points, respectively. These increases are not trivial, as they represent the difference between a C+ and a B−. Unfortunately, increased final grades are accompanied by decreased final exam scores (by 2−3 questions); thus, improved final grades may not reflect increases in student learning. However, we speculate that this decrease may be due to unfamiliarity with the multiple-choice final exam format, as has been discussed in education literature.[32] Regular exams in the traditionally graded course contained multiple-choice questions, but exams in the specifications-based course did not. There are notable limitations to the use of standardized exams as quantitative measures of learning, especially when delivered as an unfamiliar exam format.[33,34]

We observed gendered differences in performance on the ACS final exam under both grading schemes, where men outperformed women on the exam by about 3−4 questions. Though small, the difference in average final exam performance between men and women was more pronounced in the course using specifications-based grading. We speculate that this result may be due to the exam format and because high-stakes exams in STEM have been observed to be gender-biased.[35] Moreover, exams using multiple-choice questions have shown gendered results in STEM courses, with men outperforming women.[36] Given these limitations in the final exam's ability to measure learning in an unbiased way, the observed gendered decrease in the average final exam score may be less worrisome.

End-of-semester student surveys showed that specifications-based grading is associated with more positive attitudes toward student-instructor interactions, particularly those related to getting feedback, seeking out help, and asking questions. These increased interactions were particularly prevalent for women. These improvements in student perceptions contrast Noell's implementation, where students perceived specifications-based grading to be unfair and complex.[6] Importantly, our survey did not measure perceptions of the grading scheme, instead focusing on elements of the "flipped" format of the course, explaining the discrepancy.

"Flipped" courses provided ample informal opportunities for feedback (e.g., asking questions during in-class activities) for students from both grading schemes. However, based on students' anecdotes, we speculate that attitudes toward receiving feedback were more positive in the specifications-based grading regime due to the formalized "built-in" feedback (e.g., the "exam cycle" of first attempt, receive feedback, review, second attempt). According to Anderson and Carta-Falsa, students desire open and respectful learning environments created through student−instructor interactions.[37] Baepler and Walker found that in active learning courses (e.g., "flipped"

https://doi.org/10.1021/acs.jchemed.3c00473
J. Chem. Educ. 2023, 100, 4318−4326

classes) effective communication and feedback were a key dimension of interpersonal classroom relationships.[38] White et al. highlight the importance of student−instructor relationships in reducing equity gaps in STEM.[28] The implication of these studies and our results is that specifications-based grading provides formal avenues for student−instructor interactions, which may lead to more positive student attitudes toward "flipped" classrooms.

Our results also show gendered differences. While both genders' final grades improved with specifications-based grading, men's grades were slightly (not statistically significantly) higher. Men also outperformed women on the final exam. Combining these results with the observation that women were more likely to report interactions with the instructor suggests that student−instructor interactions are not the only important factor governing student outcomes. Our results do not elucidate what these factors may be.

■ LIMITATIONS

It is important to recognize the limitations of this study. As proxies for student learning, we used final letter grades and standardized exam results, which are snapshots into student learning and thus are limited in what outcomes they can report on. Another limitation is that our survey did not assess why students were more motivated to engage with the instructor or what feature of specifications-based grading led them to perceive their interactions with the instructor more positively. Thus, we can only speculate if there is connection between these results. Furthermore, we conducted this study in one course at one institution. Greater than 85% of the participants in this study were white, potentially limiting the application of these findings to more diverse settings.

Authors Gute and Wainman were the instructors of the courses included in this study. Given the gendered nature of the survey results, it is pertinent to share that both Gute and Wainman are white cisgender men. Wainman also identifies as LGBTQ+. It is possible that the differential results observed here for men and women students were influenced by whether a particular student's gender matched their instructor.[39] These limitations mean that the results may not be generalizable to other contexts.

■ LESSONS LEARNED

Despite the limitations, we are encouraged by these results and hope to see implementation of similar grading schemes. The authors have implemented the specifications-based grading scheme described above in several large-enrollment general chemistry II courses. Along the way, we learned several important lessons that helped us hone strategies to successfully address issues and concerns as they arose during the implementation of the grading scheme. We include these lessons in Supporting Information 1.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.3c00473.

Additional implementation details (PDF)
Additional implementation details (DOCX)
Additional methodology details (PDF)
Additional methodology details (DOCX)
Example materials (PDF)
Example materials (DOCX)

■ AUTHOR INFORMATION

Corresponding Author

Jacob W. Wainman − *Chemistry and Biochemistry Department, University of Minnesota Duluth, Duluth, Minnesota 55812, United States;* orcid.org/0000-0002-3829-7116; Email: jgauer@d.umn.edu

Authors

Broden Bunnell − *Chemistry and Biochemistry Department, University of Minnesota Duluth, Duluth, Minnesota 55812, United States*
Lauren LeBourgeois − *Chemistry and Biochemistry Department, University of Minnesota Duluth, Duluth, Minnesota 55812, United States*
James Doble − *Chemistry and Biochemistry Department, University of Minnesota Duluth, Duluth, Minnesota 55812, United States*
Brian Gute − *Chemistry and Biochemistry Department, University of Minnesota Duluth, Duluth, Minnesota 55812, United States;* orcid.org/0000-0003-4081-2161

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.3c00473

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Swenson Family Foundation and the Swenson College of Science and Engineering at the University of Minnesota Duluth for funding the postdoctoral fellow (J.D.) and undergraduate researchers on this project (B.B. and L.L.). We also thank the many teaching assistants who helped implement this grading scheme over the years, including Salem Baijali, Trista Betz, Anh Cong, Charlie Liggett, Faith Murphy, Austin Nickell, Maddie Petersen, and Kevin Schultz.

■ REFERENCES

(1) Harris, R. B.; Mack, M. R.; Bryant, J.; Theobald, E. J.; Freeman, S. Reducing Achievement Gaps in Undergraduate General Chemistry Could Lift Underrepresented Students into a "Hyperpersistent Zone. *Sci. Adv.* **2020**, *6* (24), No. eaaz5687.

(2) Tashiro, J.; Talanquer, V. Exploring Inequities in a Traditional and a Reformed General Chemistry Course. *J. Chem. Educ.* **2021**, 13.

(3) *American Chemical Society, Office of Professional Training*. ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs. American Chemical Society, 2015; pp 1−38. https://www.acs.org/content/dam/acsorg/about/governance/committees/training/2015-acs-guidelines-for-bachelors-degree-programs.pdf.

(4) Nilson, L. B. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*; Stylus Publising, LLC: Sterling, VA, 2014.

(5) Martin, L. J. Introducing Components of Specifications Grading to a General Chemistry I Course. In *Enhancing Retention in Introductory Chemistry Courses: Teaching Practices and Assessments*, Vol. *1330*; ACS Symposium Series; American Chemical Society: Washington, DC, 2019; pp 105−119.

(6) Noell, S. L.; Rios Buza, M.; Roth, E. B.; Young, J. L.; Drummond, M. J. A Bridge to Specifications Grading in Second Semester General Chemistry. *J. Chem. Educ.* **2023**, *100* (6), 2159−2165.

(7) Boesdorfer, S. B.; Baldwin, E.; Lieberum, K. A. Emphasizing Learning: Using Standards-Based Grading in a Large Nonmajors'

General Chemistry Survey Course. *J. Chem. Educ.* **2018**, *95* (8), 1291−1300.

(8) Toledo, S.; Dubas, J. M. A Learner-Centered Grading Method Focused on Reaching Proficiency with Course Learning Outcomes. *J. Chem. Educ.* **2017**, *94* (8), 1043−1050.

(9) Arnaud, C. H. How an Alternative Grading System Is Improving Student Learning. *Chem. Eng. News*. April 25, 2021. https://cen.acs.org/education/undergraduate-education/alternative-grading-system-improving-student/99/i15 (accessed 2023−05−20).

(10) Houseknecht, J. B.; Bates, L. K. Transition to Remote Instruction Using Hybrid Just-in-Time Teaching, Collaborative Learning, and Specifications Grading for Organic Chemistry 2. *J. Chem. Educ.* **2020**, *97* (9), 3230−3234.

(11) McKnelly, K. J.; Morris, M. A.; Mang, S. A. Redesigning a "Writing for Chemists" Course Using Specifications Grading. *J. Chem. Educ.* **2021**, *98* (4), 1201−1207.

(12) Howitz, W. J.; Frey, T.; Saluga, S. J.; Nguyen, M.; Denaro, K.; Edwards, K. D. A Specifications-Graded, Sports Drink-Themed General Chemistry Laboratory Course Using an Argument-Driven Inquiry Approach. *J. Chem. Educ.* **2023**, *100* (2), 672−680.

(13) Ring, J. ConfChem Conference on Select 2016 BCCE Presentations: Specifications Grading in the Flipped Organic Classroom. *J. Chem. Educ.* **2017**, *94* (12), 2005−2006.

(14) Donato, J. J.; Marsh, T. C. Specifications Grading Is an Effective Approach to Teaching Biochemistry. *J. Microbiol. Biol. Educ.* **2023**, *24* (2), No. e00236.

(15) Ahlberg, L. Organic Chemistry Core Competencies: Helping Students Engage Using Specifications. In *Engaging Students in Organic Chemistry*, Vol. *1378*; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; pp 25−36.

(16) Hunter, R. A.; Pompano, R. R.; Tuchler, M. F. Alternative Assessment of Active Learning. In *Active Learning in the Analytical Chemistry Curriculum*, Vol. *1409*; ACS Symposium Series; American Chemical Society: Washington, DC, 2022; pp 269−295.

(17) Hollinsed, W. C. Applying Innovations in Teaching to General Chemistry. In *Increasing Retention of Under-Represented Students in STEM through Affective and Cognitive Interventions*, Vol. *1301*; American Chemical Society: Washington, DC, 2018.

(18) Katzman, S. D.; Hurst-Kennedy, J.; Barrera, A.; Talley, J.; Javazon, E.; Diaz, M.; Anzovino, M. E. The Effect of Specifications Grading on Students' Learning and Attitudes in an Undergraduate-Level Cell Biology Course. *J. Microbiol. Biol. Educ.* **2021**, *22* (3), No. e00200.

(19) Henriksen, M.; Kotas, J.; Wentworth, M. Specifications-Based Grading Reduces Anxiety for Students of Ordinary Differential Equations. *CODEE J.* **2020**, *13* (1), 1−18.

(20) Mendez, J. *Standards-Based Specifications Grading in Thermodynamics*; Purdue University, 2018. DOI: 10.5703/1288284316862.

(21) Carlisle, S. Simple Specifications Grading. *PRIMUS* **2020**, *30* (8−10), 926−951.

(22) Prasad, P. V. Using Revision and Specifications Grading to Develop Students' Mathematical Habits of Mind. *PRIMUS* **2020**, *30* (8−10), 908−925.

(23) Williams, K. Specifications-Based Grading in an Introduction to Proofs Course. *PRIMUS* **2018**, *28* (2), 128−142.

(24) Tsoi, M. Y.; Anzovino, M. E.; Erickson, A. H. L.; Forringer, E. R.; Henary, E.; Lively, A.; Morton, M. S.; Perell-Gerson, K.; Perrine, S.; Villanueva, O.; Whitney, M.; Woodbridge, C. M. Variations in Implementation of Specifications Grading in STEM Courses. *Ga. J. Sci.* **2019**, *77*, 1−21.

(25) Ryan, M. D.; Reid, S. A. Impact of the Flipped Classroom on Student Performance and Retention: A Parallel Controlled Study in General Chemistry. *J. Chem. Educ.* **2016**, *93* (1), 13−23.

(26) Yestrebsky, C. L. Flipping the Classroom in a Large Chemistry Class-Research University Environment. *Procedia - Soc. Behav. Sci.* **2015**, *191*, 1113−1118.

(27) Gute, B. D. Building Understanding and Appreciation through Collaborative Work in General Chemistry. In *Innovative Learning and Teaching: Experiments Across the Disciplines*; University of Minnesota Libraries Publishing, 2017; pp 40−57.

(28) White, K. N.; Vincent-Layton, K.; Villarreal, B. Equitable and Inclusive Practices Designed to Reduce Equity Gaps in Undergraduate Chemistry Courses. *J. Chem. Educ.* **2021**, *98* (2), 330−339.

(29) Strayhorn, T. L. *College Students' Sense of Belonging*, 2nd ed.; Routledge: New York, 2018.

(30) Moriña, A. The Keys to Learning for University Students with Disabilities: Motivation, Emotion and Faculty-Student Relationships. *PLoS One* **2019**, *14* (5), No. e0215249.

(31) Kaur, R. Role of Faculty Mentorship in Exploring Students' Perceptions of Science Identity While Pursuing STEM Courses and Programs. *Aurora Univ. ProQuest Diss. Publ.* **2020**, n/a.

(32) Mehrens, W. A.; Popham, W. J.; Ryan, J. M. How to Prepare Students for Performance Assessments. *Educ. Meas. Issues Pract.* **1998**, *17* (1), 18−22.

(33) Sanders, W. L.; Horn, S. P. Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes. *Education Policy Analysis Archives* **1995**, *3* (6), 1−15.

(34) Gallagher, C. J. Reconciling a Tradition of Testing with a New Learning Paradigm. *Educ. Psychol. Rev.* **2003**, *15* (1), 83−99.

(35) Ballen, C. J.; Salehi, S.; Cotner, S. Exams Disadvantage Women in Introductory Biology. *PLoS One* **2017**, *12* (10), No. e0186419.

(36) Beller, M.; Gafni, N. Can Item Format (Multiple Choice vs. Open-Ended) Account for Gender Differences in Mathematics Achievement? *Sex Roles* **2000**, *42*, 1−21.

(37) Anderson, L. E.; Carta-Falsa, J. Factors That Make Faculty and Student Relationships Effective. *Coll. Teach.* **2002**, *50* (4), 134−138.

(38) Baepler, P.; Walker, J. D. Active Learning Classrooms and Educational Alliances: Changing Relationships to Improve Learning. *New Dir. Teach. Learn.* **2014**, *2014* (137), 27−40.

(39) Bettinger, E. P.; Long, B. T. Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students. *Am. Econ. Rev.* **2005**, *95* (2), 152−157.